AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions and listings of claims in the application.

Listing of Claims:

1.   (Currently Amended) A processor-implemented method of collaborative focused crawling of documents related to focus topics on a network, the method comprising:

selectively prioritizing the documents to crawl based on a set of rules;

fetching prioritized documents from the network;

for each fetched document, determining whether the fetched document is relevant to any of the focus topics;

crawling the fetched document that matches any of the focus topics such that the fetched document is crawled only once even if the fetched document matches a plurality of the focus topics, wherein the fetched document comprises a document of interest for access by a user;

further crawling out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest;

determining whether the fetched document should be disallowed, and upon determination that the fetched document should be disallowed:

selectively disallowing the fetched document;

identifying a resource locator string associated with the disallowed fetched document; and

placing the resource locator string for the disallowed fetched document in a blacklist in order to prevent future crawling of the fetched document;

wherein the crawling is performed using a collaborative focus by analyzing the documents for more than one focus topic of the focus topics at a

time.

2.   (Original) The method of claim 1, further comprising seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents.

3.   (Original) The method of claim 2, further comprising crawling the seed uniform resource locator strings.

4.   (Original) The method of claim 3, further comprising writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings.

5.   (Original) The method of claim 4, further comprising a foreman function for reading a plurality of contents of the resulting uniform resource locator strings.

6.   (Original) The method of claim 5, further comprising the foreman function passing the contents of the resulting uniform resource locator strings to a miner.

7.   (Original) The method of claim 6, further comprising the miner instructing a fetcher to crawl a plurality of out-links on a document of the resulting resource locator string when the contents of the resulting resource locator string match a focus topic of the miner.

8.   (Original) The method of claim 6, further comprising the miner ignoring resulting resource locator string when the contents of the resulting resource locator string do not match the focus of the miner.

9.   (Original) The method of claim 6, further comprising the miner managing a plurality of focus topics.

10.   (Original) The method of claim 9, further comprising the miner allowing a crawling of the resulting resource locator string when the resulting resource locator string matches a plurality of web space rules.

11.   (Previously Presented) The method of claim 10 , wherein the web space rules comprise domain rules, IP address rules, and prefix rules.

12.   (Previously Presented) The method of claim 10, further comprising the miner disallowing the crawling of the resulting resource locator string when the content of the resulting resource locator string matches a focus topic of the miner.

13.   (Previously Presented) The method of claim 10, wherein the miner comprises an unfocus miner that places the resulting uniform resource locator strings that match an unfocus topic in the blacklist, so that the uniform resource locator strings will not be crawled again.

14.   (Currently Amended) A computer program product having a plurality of executable instruction codes stored on a computer readable storage medium, for implementing a collaborative focused crawling of documents related to focus topics on a network, the computer program product comprising:

        a first set of instruction codes for selectively prioritizing the documents to crawl based on a set of rules;

        a second set of instruction codes for fetching prioritized documents from the network;

        for each fetched document, a third set of instruction codes determines whether the fetched document is relevant to any of the focus topics;

a fourth set of instruction codes for crawling the fetched document that matches any of the focus topics such that the fetched document is crawled only once even if the fetched document matches a plurality of the focus topics, wherein the fetched document comprises a document of interest for access by a user;

wherein the fourth set of instruction codes further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest;

wherein the fourth set of instruction codes further determine whether the fetched document should be disallowed, and upon determination that the fetched document should be disallowed;

selectively disallowing the fetched document;

identifying a resource locator string associated with the disallowed fetched document; and

placing the resource locator string for the disallowed fetched document in a blacklist in order to prevent future crawling of the fetched document;

wherein the crawling is performed using a collaborative focus by analyzing the documents for more than one focus topic of the focus topics at a time.

15.   (Original) The computer program product of claim 14, further comprising a fifth set of instruction codes for seeding a plurality of seed uniform resource locator strings to start the collaborative focused crawling of the documents.

16.   (Original) The computer program product of claim 15, wherein the fourth set of instruction codes further crawls the seed uniform resource locator strings.

17.   (Original) The computer program product of claim 16, further comprising a sixth set of instruction codes for writing a plurality of resulting uniform resource locator strings obtained by crawling the seed uniform resource locator strings.

18.   (Currently Amended) A processor-implemented system for implementing a collaborative focused crawling of documents related to focus topics on a network, the system comprising:

an evaluator that selectively prioritizes the documents to crawl based on a set of rules;

a fetcher that fetches prioritized documents from the network;

for each fetched document, a focus engine determines whether the fetched document is relevant to any of the focus topics;

a crawler for crawling the fetched document that matches any of the multiple focus topics such that the fetched document is crawled only once even if the fetched document matches a plurality of the focus topics, wherein the fetched document comprises a document of interest for access by a user;

wherein the crawler further crawls out-links on the fetched document based on an assumption that if the fetched document is of interest, the out-links are also of interest;

wherein the crawler further determines whether the fetched document should be disallowed, and upon determination that the fetched document should be disallowed:

selectively disallowing the fetched document;

identifying a resource locator string associated with the disallowed fetched document; and

placing the resource locator string for the disallowed fetched document in a blacklist in order to prevent future crawling of the fetched document;

wherein the crawler uses a collaborative focus by analyzing the documents for more than one focus topic of the focus topics at a time.

.

19. (Previously Presented) The system of claim 18, further comprising a plurality of seed uniform resource locator strings that are used to initiate the collaborative focused crawling of the documents.

20. (Previously Presented) The system of claim 19, wherein the crawler further crawls the seed uniform resource locator strings.